Empirical model of savings

Previous CNA studies have modeled and predicted Navy A-76 savings directly [2]. These studies estimate a regression model in which the dependent variable is savings from completed A-76 competitions. Based on the model estimates, predicted savings for each function in the inventory can be obtained directly. The approach does not require an equation to predict baseline cost or who will win the competition as would a model predicting percentage savings or predicting individual bids.

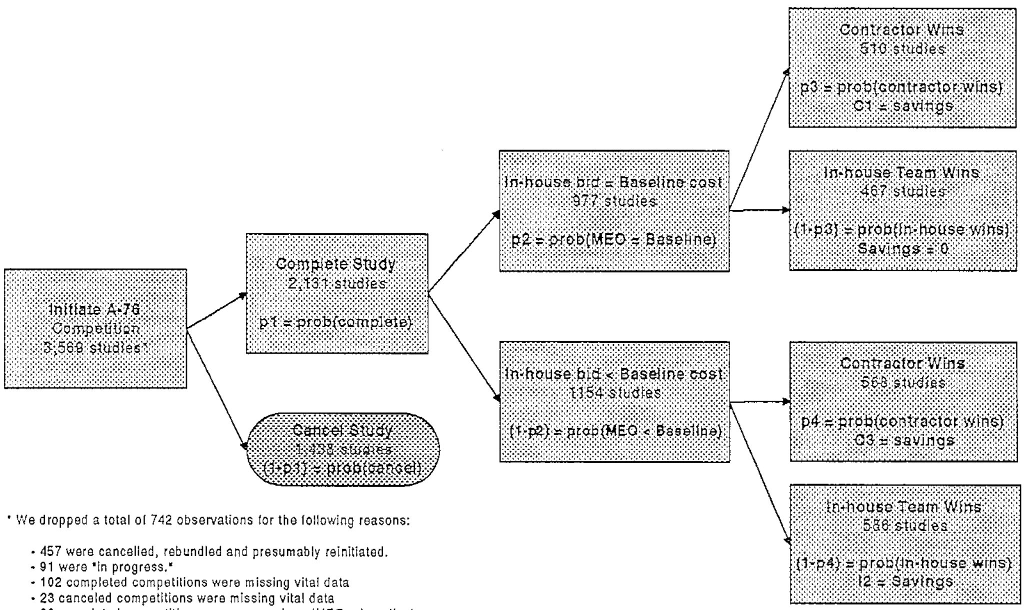
In addition, since the ordinary least squares (OLS) regression goes through the means of the explanatory variables, it will produce a positive predicted savings for all those functions in the inventory that have values for the independent variables close to the means. In this paper, we chose to expand upon the previous regression model of predicting savings as explained below.

A sequential decision tree of A-76 savings

The discussion associated with figure 1 suggests modeling the A-76 process as a sequential decision tree with different expected savings depending on which branch a particular study follows. Given the probability of following each branch, overall expected savings can be calculated as the probability of following each branch times the expected savings from following that branch. This is the approach used here and shown in figure 4.[10]

10. Our other models will be documented later.

Figure 4. Empirical model of A-76 competition savings



* We dropped a total of 742 observations for the following reasons:

- 457 were cancelled, rebundled and presumably reinitiated.
- 91 were "in progress."
- 102 completed competitions were missing vital data
- 23 canceled competitions were missing vital data
- 60 completed competitions were expansions (MEO > baseline)
- 5 canceled competitions were expansions (MEO > baseline)
- 3 had the contractor winning with a higher bid
- 1 was an apparent typographical error

$$E(\text{savings}|\text{complete}) = p_2 \cdot p_3 \cdot C_1 + (1 - p_2) \cdot (1 - p_4) \cdot I_2 + (1 - p_2) \cdot p_4 \cdot C_3$$

The process that ultimately produces observed savings from A-76 competitions involves several steps which we will discuss individually.

The probability of completion

Given that a CA study has been initiated and the scope of work was not increased, the next step is a decision on whether or not to complete the study.[11] Define indicator variable $Y_{1,i}$ such that:

11. In 65 of the 4,311 functions studied since 1979, the scope of work was increased so that the MEO was greater than the baseline number of billets. For these rare cases, it is difficult to define baseline cost since the job being competed has not been previously performed by an in-house team. For this reason, we will not include these cases in the analysis.

$Y_{1,i} = 1$ if the study i is completed

$Y_{1,i} = 0$ if the study i is not completed.

Estimating the probability of completion

The determinants of $Y_{1,i}$ can be estimated with a probit model.[12] Let the probability that $Y_{1,i} = 1$ for function i be denoted by $p_{1,i}$. In the probit model, $p_{1,i}$ is a nonlinear function of a vector of exogenous variables $X_{1,i}$ with the constraint $0 < p_{1,i} < 1$ being imposed on the functional form.

The probit model assumes we have a regression model given by

$$Y_{1,i}^* = \beta'_1 X_{1,i} + u_{1,i} , \tag{1}$$

where $u_{1,i}$ is a normally distributed random variable with zero mean and unit variance, β_1 is a vector of unknown parameters to be estimated, and $Y_{1,i}^*$ is not observed. It is common to refer to the variable $Y_{1,i}^*$ as a "latent" variable. What is observed is a dummy indicator variable $Y_{1,i}$ defined by

$$Y_{1,i} = 1 \text{ if } Y_{1,i}^* > 0 \tag{2}$$

$$Y_{1,i} = 0 \text{ otherwise} .$$

From equations (1) and (2), we obtain:

$$\begin{aligned} P_{1,i} &= Prob(Y_{1,i} = 1) = Prob(u_{1,i} > -\beta'_1 X_{1,i}) \\ &= 1 - F(-\beta'_1 X_{1,i}) = F(\beta'_1 X_{1,i}) , \end{aligned} \tag{3}$$

where $F(\cdot)$ is the standard normal cumulative distribution function. The unknown parameters in the matrix β_1 can be estimated with a maximum likelihood procedure.

12. See [16 through 20] for a discussion of the probit model and its extensions.

The probability of the in-house bid equaling the baseline costs

For those studies that are completed, the in-house team must decide whether or not to keep the Most Efficient Organization (MEO), which is a major determinant of the in-house bid, at the current number of billets or to lower it. Define the indicator variable $Y_{2,i}$ such that

$Y_{2,i} = 1$ if the in-house bid equals the baseline cost

$Y_{2,i} = 0$ otherwise .

Let the probability that $Y_{2,i} = 1$ for function i be denoted by $p_{2,i}$. If the determinants of $p_{2,i}$ depend on a vector of exogenous variables $X_{2,i}$, a model similar to equation (3) will produce estimates of the unknown parameters in β_2.

The probability of the contractor winning

When $Y_{2,i} = 1$, either the in-house team or the outside contractor is awarded the job. If the in-house team wins, savings will be zero. If the contractor wins, there will be positive savings. Define the indicator variable $Y_{3,i}$ such that

$Y_{3,i} = 1$ if the contractor wins given MEO = baseline

$Y_{3,i} = 0$ if the in-house team wins given MEO = baseline.

Denote the probability that $Y_{3,i} = 1$ as $p_{3,i}$. The unknown parameters β_3 can be estimated with a probit model.

Finally, when $Y_{2,i} = 0$, either the in-house team or the outside contractor is awarded the job. Define the indicator variable $Y_{4,i}$ such that

$Y_{4,i} = 1$ if the contractor wins given MEO < baseline

$Y_{4,i} = 0$ if the in-house team wins given MEO < baseline.

Denote the probability that $Y_{4,i} = 1$ as $p_{4,i}$. The unknown parameters β_4 can be estimated with a probit model.

Expected savings

The above four indicator variables define a sequential decision tree that generates observations on savings from A-76 competitions. There are four different ways to generate observed savings. These are:

1. $Y_{1,i} = 1$, $Y_{2,i} = 1$ and $Y_{3,i} = 0$.

In this case, the in-house team decides to keep the number of billets in its bid at the current level, and it wins the contract. Therefore, observed savings will be zero.

2. $Y_{1,i} = 1$, $Y_{2,i} = 1$ and $Y_{3,i} = 1$.

In this case, the in-house team's bid for the MEO is the same as the current level, but the contractor wins the contract. The observed savings will be positive. Denote these savings as $C_{1,i}$ and assume they are related to a vector of exogenous variables, $Z_{1,i}$, via the stochastic regression equation

$$C_{1,i} = \gamma'_1 Z_{1,i} + \varepsilon_{1,i}, \tag{4}$$

where $\varepsilon_{1,i}$ is a random error term with zero mean and γ_1 is vector of parameters that can be estimated with ordinary least squares on the observed savings.

3. $Y_{1,i} = 1$, $Y_{2,i} = 0$ and $Y_{3,i} = 0$.

In this case, the in-house team's bid contains fewer billets than the current level, and it wins the contract. Again, the observed savings will be positive. Denote these savings by $I_{2,i}$ and assume they are related to a vector of exogenous variables $Z_{2,i}$ via the stochastic regression equation

$$I_{2,j} = \gamma'_2 Z_{2,i} + \varepsilon_{2,j}, \tag{5}$$

where γ_2 is vector of parameters that can be estimated with ordinary least squares on the observed savings.

4. $Y_{1,i} = 1$, $Y_{2,i} = 0$ and $Y_{3,i} = 1$.

In this case, the in-house team's bid contains fewer billets than the current level, but the contractor wins the contract. Again, the observed savings will be positive. Denote these savings by $C_{3,i}$ and assume they are related to a vector of exogenous variables $Z_{3,i}$ via the stochastic regression equation

$$C_{3,\,i} = \gamma'_3 Z_{3,\,i} + \varepsilon_{3,\,i} \quad , \tag{6}$$

where γ_3 is vector of parameters that can be estimated with ordinary least squares on the observed savings.

The decision tree that generates the observations on A–76 competition savings is depicted in figure 4. Let S_i be the savings associated with study i. The expected savings for function i drawn randomly from the inventory is

$$E(S_i) = p_{1,i}p_{2,i}p_{3,i}E[\,C_{1,i}] + p_{1,i}(1\text{-}p_{2,i})\,(1\text{-}p_{4,i})\,E[I_{2,i}] + \\ p_{1,i}\,(1 - p_{2,i})p_{4,i}E[\,C_{3,i}], \tag{7}$$

where $E[\,C_{1,i}]$ is the expected savings for function i if the MEO equals the current billets and the contractor wins the bid, $E[\mathbf{I}_{2,i}]$ is the expected savings if the MEO is less than the current billets and the in-house team wins the bid, and $E[\,C_{3,i}]$ is the expected savings when the MEO is less than the current billets and the contractor wins the bid.

Equation 7 assumes only the proportion $p_{1,i}$ of all A–76 competitions will be completed. If one is interested in computing the potential expected savings assuming all functions in the inventory are studied and completed, this expectation is given by

$$E(\,S_i \mid \text{i } \textbf{completed}) = p_{2,i}p_{3,i}E[\,C_{1,i}] + (1\text{-}p_{2,i})\,(1\text{-}p_{4,i})E[I_{2,i}] + \\ (1\,\text{-}p_{2,i})p_{4,i}E[\,C_{3,i}]\;, \tag{8}$$

where all variables are as previously defined.